MPG analysis, automatic or manual?

Tom van Dienst

Executive summary

Taking several attributes in account; MPG can be affected by the weight and the transmission of the car. A higher intercept and steeper slope shows that a manual transmission is only beneficial for lighter weight cars. Heavier cars benefit from an automatic transmission with a lower intercept and lighter slope.

Main analysis

Data exploration

We make a small alteration to our dataset. The standard variables of am are 0 (automatic) and 1(manual), we'll rename those to their actual descriptions to prevent confusion. (See Appendix, Datsets) Let's take a look at the mpg for automatic vs manual transmission. (See Appendix, Figure 1)

It's clear that if you just look at the transmission, an automatic transmission has a much lower MPG than a manual one. But of course there are other attributes which might have something to do with this. Let's take a look at the complete set. We'll see the various measures vs mpg with the transmission as a color. We'll also include a regression line for these variables. (See appendix A, Figure 2)

So what do we see? first of all we see that most of the heavier vehicles tend to have an automatic transmission. This seems to match with the displacement and the number of cylinders. Displacement is the total volume of the cylinders, so this is not very surprising. A higher displacement, and weight have a negative influence on the MPG. One value with outliers is the HP, it seems there are some manuals that have a higher horsepower than automatics. This doesn't always match up with the weight. It seems there are some cars with a low weight but high horsepower. One more attribute of interest is 1/4 mile time (qsec), it seems they follow the same slope but have a completely different intercept.

Model building

So what will we assume for this model? Our analysis shows that mpg is influenced by weight, displacement, horsepower and qsec. Weight and displacement have a correlation of 0.89 and is considered highly related so it's enough to just take weight. We'll set up several models and run them through ANOVA to see what influences we can find. We'll start with weight since it has such a big impact on the mpg, next we'll add hp and qsec.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt * am
## Model 2: mpg ~ wt * am + hp * am
## Model 3: mpg ~ wt * am + hp * am + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     28 188.01
## 2     26 135.90  2    52.108 5.6915 0.009184 **
## 3     25 114.44  1    21.459 4.6877 0.040126 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the analysis shows, a combination of all three is most succesful. Every addition of an attribute decreases our variance significantly.

Let's take a look at the residuals in our model (See Appendix, figure 3)

The remaining variance seems to be evenly distributed with one outlier in the top. Let's take a look at the output when we remove/add elements with the dffits function:

```
mtdatashow <- mtdata
mtdatashow$fits <-dffits(fitwthpqsec)
kable(mtdatashow[abs(mtdatashow$fits)>1, c("mpg", "wt", "hp", "qsec", "fits")])
```

	mpg	wt	hp	qsec	fits
8	24.4	3.190	62	20.00	1.180166
17	14.7	5.345	230	17.42	1.086786
18	32.4	2.200	66	19.47	1.236905
31	15.0	3.570	335	14.60	1.281352

There are only 4 cars who have more than 1 mpg impact when removed from the equation. The biggest culprit is a car with low weight and high horsepower. One of the outliers we saw during our data exploration.

What are our coefficients and confidence intervals?

	Estimate	Pr(>|t|)	2.5 %	97.5 %	Description
(Intercept)	12.225	0.181	-6.093	30.544	Intercept for automatic
wt	-2.512	0.013	-4.437	-0.586	Weight
amManual	14.876	0.001	6.665	23.088	Intercept difference for manual
hp	-0.011	0.553	-0.05	0.027	Horsepower
qsec	0.89	0.04	0.043	1.738	1/4 mile time
wt:amManual	-5.212	0.013	-9.24	-1.183	Weight covariate difference for manual
amManual:hp	0.015	0.441	-0.024	0.054	HP covariate difference for manual

Our confidence intervals show that the Intercept (both for automatic and the difference with manual) and the covariates wt(automatic and manual difference) and qsec will have a consistent influence (either negative or positive) with 95% certainty.

Conclusion

There's definitely a difference between automatic and manual. The model for manual cars shows that they start with a higher initial MPG. But (with all other variables remaining equal) an increase in 1000 lbs for an automatic results in a 2.5 decrease in MPG. The same increase in weight for a manual car would result in $2.5 + 5.2 = 7.7$ decrease in MPG. The other variable of interest is the 1/4 mile time, but this variable does not have a seperate covariate for manual cars and applies for both transmissions. These covariates all have confidence intervals that do not contain 0, and are therefore significant enough to be mentioned. Other variables do not have enough effect.

So what can we conclude? If you're driving a small car, you're beter off with a manual transmission. But when you're driving a heavy car, you're better off switching to an automatic. Keep in mind that a higher 1/4 mile time will also have a slightly positive influence on your mpg.

Appendix A

Original files

Because of the limit imposed on the number of pages, I've made the original files available in a github repository: **https://github.com/Tothalvadi/Regression-Models** The code can be found in the rmd file.

Dataset

Below the alterations I made to the original dataset. 0 values in AM are translated to "Automatic" and other values to "Manual"

```
mtdata  <- mutate(mtcars
                  , am = factor(ifelse(am ==  0, "Automatic", "Manual"))
)
```

Figure 1

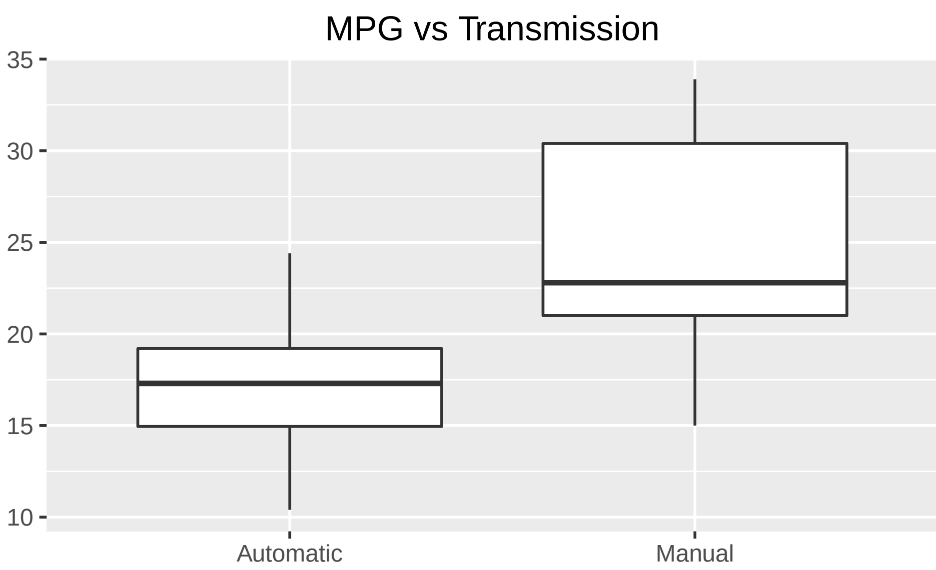

Figure 2

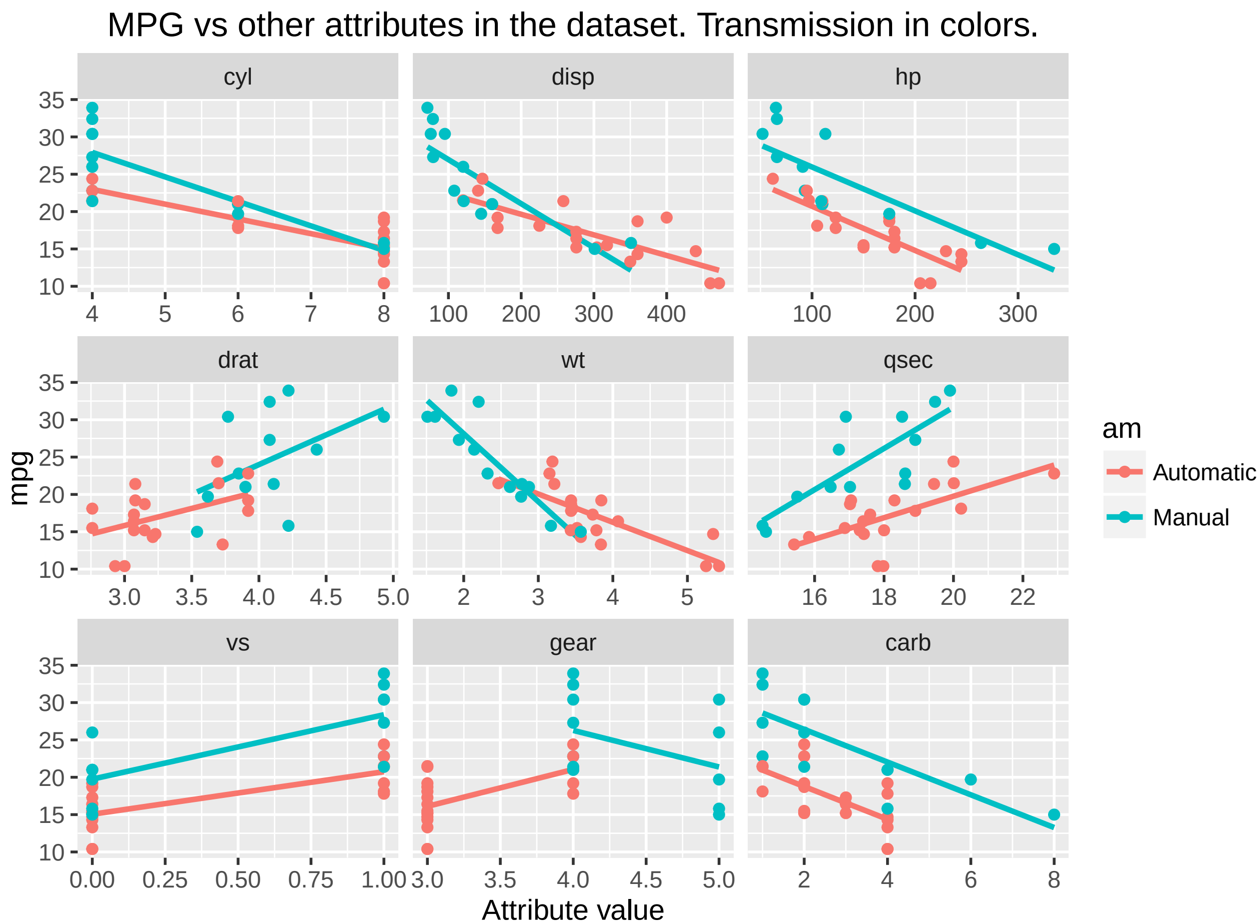

Figure 3